

Article

Evaluating a Taxonomy of Textual Uncertainty for Collaborative Visualisation in the Digital Humanities

Alejandro Benito-Santos ^{1,*}, Michelle Doran ², Aleyda Rocha ³, Eveline Wandl-Vogt ^{3,4,5}, Jennifer Edmond ²
and Roberto Therón ¹

¹ VisUSAL Research Group, Universidad de Salamanca, 37004 Salamanca, Spain; theron@usal.es

² Trinity Long Room Hub Arts & Humanities Research Institute, University of Dublin Trinity College, Dublin 2, Ireland; doranm1@tcd.ie (M.D.); edmondj@tcd.ie (J.E.)

³ Austrian Centre for Digital Humanities and Cultural Heritage, Austrian Academy of Sciences, 1010 Vienna, Austria; aleyda.rocha@oeaw.ac.at (A.R.); eveline-wandl-vogt@oeaw.ac.at (E.W.-V.)

⁴ Ars Electronica Research Institute “Knowledge for Humanity”, Ars Electronica, 4040 Linz, Austria

⁵ MetaLab (at) Harvard, Cambridge, MA 02138, USA

* Correspondence: abenito@usal.es

Abstract: The capture, modelling and visualisation of uncertainty has become a hot topic in many areas of science, such as the digital humanities (DH). Fuelled by critical voices among the DH community, DH scholars are becoming more aware of the intrinsic advantages that incorporating the notion of uncertainty into their workflows may bring. Additionally, the increasing availability of ubiquitous, web-based technologies has given rise to many collaborative tools that aim to support DH scholars in performing remote work alongside distant peers from other parts of the world. In this context, this paper describes two user studies seeking to evaluate a taxonomy of textual uncertainty aimed at enabling remote collaborations on digital humanities (DH) research objects in a digital medium. Our study focuses on the task of free annotation of uncertainty in texts in two different scenarios, seeking to establish the requirements of the underlying data and uncertainty models that would be needed to implement a hypothetical collaborative annotation system (CAS) that uses information visualisation and visual analytics techniques to leverage the cognitive effort implied by these tasks. To identify user needs and other requirements, we held two user-driven design experiences with DH experts and lay users, focusing on the annotation of uncertainty in historical recipes and literary texts. The lessons learned from these experiments are gathered in a series of insights and observations on how these different user groups collaborated to adapt an uncertainty taxonomy to solve the proposed exercises. Furthermore, we extract a series of recommendations and future lines of work that we share with the community in an attempt to establish a common agenda of DH research that focuses on collaboration around the idea of uncertainty.

Keywords: human–computer interaction; digital humanities; cultural heritage; information visualisation; data modelling; uncertainty taxonomy; decision-making support; user studies



Citation: Benito-Santos, A.; Doran, M.; Rocha, A.; Wandl-Vogt, E.; Edmond, J.; Therón, R. Evaluating a Taxonomy of Textual Uncertainty for Collaborative Visualisation in the Digital Humanities. *Information* **2021**, *12*, 436. <https://doi.org/10.3390/info12110436>

Academic Editor: Stefan Jänicke

Received: 15 July 2021

Accepted: 14 October 2021

Published: 21 October 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The adequate capture and communication of uncertainty in computational processes is a long-standing challenge of computer science and related fields, such as information visualisation (IV) or the digital humanities (DH), while the introduction of uncertainty in the data analysis pipeline is known to greatly increase its complexity, it also has many proven benefits in enabling users to make better-informed decisions based on data. For this reason, there is the question of how much uncertainty needs to be incorporated into such systems to make them useful while keeping them manageable and accessible at the same time, which typically depends on the task, users’ level of expertise, or degree of familiarity with the data, among other traits.

Beyond guiding decision making, the DH community also sees representing uncertainty as an opportunity to enable many different personal *interpretations* of data, producing what the visual theory scholar J. Drucker coins as “*capta*”, which is a term conceived to stress the constructivist stance humanities scholars adopt to approach knowledge. This is in contrast to the term “data”, which refers to knowledge that is a “given” that can be “recorded and observed” [1]. Thus, “*capta*” makes reference to information that is consumed and *interpreted* by a human actor, constituting a *perspective*. This differentiation is especially important in the context of DH research, in which typically, there is a lack of ground truth personal claims can be compared to and, rather, the confrontation of multiple points of view or *perspectives* [2] is sought after, setting up an interesting scenario for collaboration centred around the uncertainty that is perceived by each participant.

Fuelled by Drucker’s ideas, in recent times, digital humanists and visualisation scholars have demonstrated an increasing interest in creating models to assess, quantify, and classify uncertainty in digital research objects in a wide variety of disciplines [3]. An example of this interest is the recent Special Issue (SI) “Uncertainty in Digital Humanities” that some of us edited in 2019. The SI accepted five contributions from distinct groups of DH experts focusing, for example, on the visualisation of digital cultural collections [2], uncertainty modelling applied to archaeological and historical research [4], or the development of digital research tools [5], among others. For example, in their paper, Windhager et al. [2] present multiple uncertainty representation techniques that can be introduced in a multi-perspective digital research environment.

Traditionally, annotation has been considered a form of *active reading* that reflects the readers’ thinking processes, effectively enhancing memorisation and reasoning over a corpus to support problem-solving. Annotations can be employed later by the same or another reader to conduct “augmented” reading, possibly leading them to formulate new hypotheses about the contents of the text. For this reason, annotation is a fundamental scholarly practice in text-oriented DH, as it is employed by scholars and students alike to perform all sorts of analyses (e.g., lexicographical, grammatical, semantic) on a given corpus. As such, digital annotation is the logical evolution of the traditional paper-based annotation and brings many advantages over the latter, such as rich text annotation, or content and media linking (e.g., of text fragments, URLs, or images), that can enhance the reading and learning experience.

Another more recent paradigm centred around the annotation task, the collaborative annotation system (CAS) [6], has emerged along with the rise of the WWW and ubiquitous computing systems. A CAS “allows users to add valued information, share ideas and create knowledge” in a digital environment. Although several examples of CAS can be found in the literature [6–9], we can identify three main shortcomings of these works that inspired us to perform the work presented in this paper: first, existing approaches do not adopt specific strategies to communicate uncertainty, which is important in a DH research context. We partially attribute this to the fact that these systems do not incorporate specific models for uncertainty, making it hard to capture, store, and communicate it. Second, the majority of the existing solutions were not designed with the nuances and goals of DH research in mind, which means it is often hard for DH scholars to adapt them to their particular needs. For example, many of these systems do not support texts given in TEI format, which is a de facto standard in DH. Therefore, scholars often are forced to convert their TEI texts into other intermediate compatible formats to be able to use these platforms. Finally, another common drawback of these platforms is that they do not make use of advanced text and information visualisation techniques that we believe could be highly beneficial in this context, attending to recent collaborations between visualisation practitioners and digital humanists. In this regard, the field of visualisation for the digital humanities (VIS4DH) has become a promising interdisciplinary field that is enabling unprecedented collaborations between engineers, designers, and humanists [10,11], bringing great benefits to both sides. Moreover, the management and display of uncertainty has turned into a growing point of interest in the field of information visualisation [12]. However, the current

literature still lacks examples of *applied* uncertainty visualisation research [13], calling for novel interdisciplinary endeavours and related user-based evaluations for which the DH represent a highly interesting field of experimentation [3]. Thus, in this context, we identify an opportunity for proposals of CAS that embrace the idea of *collaborative visualisation* to leverage the inherent cognitive load and complexity of the annotation and close/distant reading tasks.

The term *collaborative visualisation* refers to “the shared use of computer-supported, (interactive,) visual representations of data by more than one person with the common goal of contribution to joint information processing activities” [14]. Under this conception, our aim in this study is to establish the data model that a system of these characteristics would employ. To this end, we examine the task of collaborative uncertainty annotation in DH research through two workshops set in the context of the PROVIDEDH project (<https://providedh.eu/>, accessed on 13 October 2021). Here, we seek to unveil new modes of collaborative uncertainty visualisation that are first enabled in the annotation task. In general, the PROVIDEDH project seeks to enable a digital space for DH scholars to collaboratively explore and assess the evolution of uncertainty in digital research objects by visual means, allowing them to share their perspectives and insights with other stakeholders using an interdisciplinary approach. To achieve these goals, the project relies on the annotation of digital texts such as TEI (<https://tei-c.org/>, accessed on 13 October 2021) documents, which is typical of DH scholarship [15]. To this end, the implementation of knowledge regarding co-creation interactions and the outcomes of participatory design derived from this work not only contributes to the understanding of “uncertainty” in the DH domain, but also sheds new light on novel approaches to DH scholarship centred around the idea of uncertainty in the annotation task. Beyond that, our contribution informs a participatory research-based process for designing the foundations of visual text, collaborative annotation systems centred around the idea of uncertainty in the DH research process that can be used by other fellow researchers who may be involved in similar research.

2. Related Work

Our work is largely inspired by the past contributions to the fields of uncertainty modelling, visualisation design, and citizen science that are presented below.

2.1. Uncertainty Taxonomies

Due to its central role in this work, we dedicate this section to summarising and discussing the motivations behind the adoption of an uncertainty taxonomy for the DH, some of which can be found in two of our previous papers [16,17].

The creation of a taxonomy for visualising uncertainty in text corpora has been one of our main lines of research in the PROVIDEDH project since its inception in 2018. In order to design effective systems that help analysts to make decisions based on data, it is key to reflect on the notion and implications of uncertainty itself. Identifying the stages of the analysis pipeline in which this uncertainty can be introduced is of vital importance for the conception of data structures, algorithms and other mechanisms that may allow its final representation in a collaborative visualisation environment. In this regard, over the last few decades, substantial effort has been put into providing theoretical models for uncertainty management. The categorisation and assessment of uncertainty have produced many academic contributions from different areas of human knowledge, ranging from statistics and logic to philosophy or to computer science. Drawing from its parent body of research, cartography, geographical information science (GIScience: https://en.wikipedia.org/wiki/Geographic_information_science, accessed on 13 October 2021) scholars typically have a special interest in producing uncertainty typologies and taxonomies in all their forms. Carefully presenting uncertain information in digital maps is key for analysts to make more informed decisions on critical tasks such as storm and flood control, census operations, and the categorisation of soil and crop. Concretely, our taxonomy was inspired by the notable

contributions to GIScience and decision-making by Fisher [18] (Figure 1) and Smithson [19]. In their works, both authors proposed categorisations of uncertainty that attended to the *quality* of the objects' definitions to divide them into different categories. For example, Fisher distinguished between well and poorly defined objects, linking each type to aleatoric (or "irreducible") and epistemic ("reducible") uncertainty, respectively. More recently, Simon et al. [20] employed this classification in their work on the numerical assessment of risk, further dividing epistemic uncertainty into another four categories: imprecision, ignorance, credibility, and incompleteness. Our work focuses on these manifestations of epistemic uncertainty, because by definition it can be reduced, and thus resolved, in a collaborative manner. Below, we provide short definitions for each category. For a more in-depth explanation, we refer the reader to Therón et al. [17].

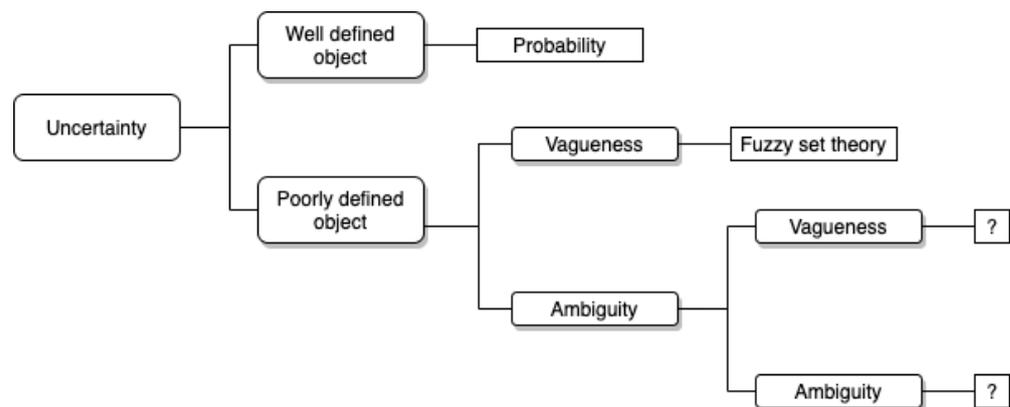


Figure 1. Fisher et al.'s taxonomy of uncertainty. Our work focuses on categories of poorly defined objects [20].

1. **Aleatoric:** Aleatoric uncertainty (also known as *statistical* or *stochastic* uncertainty) exists due to the random nature of physical events, and it is probabilistic in nature. Its main characteristic is that it is irreducible and thus it needs to be tackled using statistical artefacts. Aleatoric uncertainty is typically modelled with a continuous variable, and thus it is best represented by a continuous probability distribution function (CDF). In our recent work, we established that in the context of digital humanities, aleatoric uncertainty can be better understood as algorithmic uncertainty (e.g., in topic modelling, the probability of a word to be in a set of topics). We did not consider this type of uncertainty in this study.
2. **Epistemic:** This type of uncertainty results from a lack of knowledge and it is associated with the user performing the analysis. As opposed to aleatoric uncertainty, epistemic uncertainty can be resolved if more knowledge is gathered. Furthermore, it is known in the literature as systematic, subjective, or reducible uncertainty. This is the type of uncertainty we address in this study. In this study, we aimed to capture human judgements on epistemic uncertainty. Since humans cannot emit judgements with a high level of precision (i.e., nobody says that they are 2.3575% "sure" about something), we employed a 5-point Likert scale to capture these statements, as is common practice in the literature [21]. Thus, the logical step is to employ a discrete probability distribution to represent it [22].
 - (a) **Imprecision:** Refers to the observer's inability to pinpoint the exact value of a measure due to a lack of information (e.g., the observations made with a cheap microscope are more imprecise than those we would obtain with a better one). For example, in our context, this kind of uncertainty is often seen as imprecise statements referring to time periods (e.g., 1095–1291, the first half of 15th century, etc.).
 - (b) **Ignorance:** Ignorance is related to the fact that information can be incorrectly processed by the individuals analysing it. It is a measure of the individual's

- degree of confidence in their judgement, for example, when trying to interpret an old piece of text due to not being able to determine the meaning of a certain word, or lacking the necessary historical knowledge to understand a passage.
- (c) **Credibility:** This is also known as discord in the literature. This type of uncertainty is linked to the influence of personal bias when making a judgement, which can result in disparate (and sometimes even opposing) points of view from different agents when observing the same data. Credibility is also closely related to the concept of *authority*: for example, data that have been curated by an expert in the field are presumed to contain fewer errors by an observer. However, the concept of expertise is deeply rooted in the observer's own previous knowledge and biases, which will lead them to (consciously or unconsciously) adjust the weight given to these judgements in their decision-making process.
- (d) **Incompleteness:** Incompleteness can be regarded as a special type of imprecision, in which it becomes impossible to know every single aspect of an event. An important point is that in this case, the agent performing the analysis is fully aware of what the data should look like to be complete (i.e., they know what the complete data should look like). Take, for example, the case of a scholar working with an old registry of people's names. The first lines of the registry contain a summary stating that there are 1000 names listed. However, only a few pages were preserved, while others were lost, and only half of the names (500) are shown. The registry is imprecise because its exact original content cannot be retrieved, but also it is incomplete because the scholar knows it has missing information.

In this work, we present two user-driven experiences that sought to evaluate this taxonomy in the context of computer-supported collaborative annotation in the DH domain, in an attempt to understand how DH scholars and users can employ it to support collective reasoning in real-world analysis scenarios by setting the underlying uncertainty and data models that would govern the implementation of a visual CAS.

2.2. Uncertainty in TEI

The TEI format permits the tagging of uncertainty in digital texts using the `<certainty>` tag and the `@cert` tag attribute (<https://www.tei-c.org/Vault/P5/4.1.0/doc/tei-p5-doc/en/html/CE.html>, accessed on 13 October 2021). However, this feature is often not used. In a recent paper by us [23], we argue the reason behind this underuse is due to these two instruments not being able to cover complex cases of human-made uncertainty identification and markup in the annotation task (i.e., the annotation of epistemic uncertainty), which is in line with previous work by Binder et al. [24]. In their work on uncertainty encoding in literary annotations, the authors stress the inadequacy of the `@cert` attribute to communicate uncertainty to human users, which is instead more suitable for "indicating possible vagueness in a machine-readable way". Binder and their colleagues resolved this issue by allowing the addition of `<note>` elements to the running text, which were to be used by the editors "for supplying small texts that will be presented to the user describing the kind of uncertainty involved" [24]. However, their solution fell short in providing a formal structure for accommodating an uncertainty taxonomy like ours, which made it inadequate for our purposes.

Therefore, and inspired by Binder et al., we built our own solution as a backwards-compatible re-implementation of the `<certainty>` tag that allows the assignment of subjective categorical values (i.e., a Likert scale) via the `@cert` attribute [23]. In addition, it also supports the use of the `@degree` attribute to employ numerical values of uncertainty to cover for cases of automatic annotation (i.e., aleatoric or algorithmic uncertainty). This modification was implemented and tested in PROVIDEDH's digital platform, and it was used to drive the annotation tasks of the workshops presented in this paper. However, we

are currently working towards making this modification part of the TEI standard, and we hope it will soon become available to the rest of the community.

2.3. Collaborative Visualisation

Previously, we have referred to collaborative visualisation as an important area of computer-supported cooperative work (CSCW) in which we place the work presented in this paper. A first formal definition of the field was given by Isenberg et al. in 2011 [14], in which the authors divided this kind of approach into two major research streams according to their main kind of spatial collaboration scenario: distributed and co-located. In the first type of collaboration, distributed collaborative systems allow distant peers to perform remote work on large scientific datasets, which is typically supported by web-based environments [2,25]. In the second setting, peers are situated next to each other and collaboration is enabled, for example, via tabletop or multi-screen displays [26]. In our work, the exercises that were given to the participants during the workshops were performed in a co-located manner, while our aim is to use the knowledge gathered during these sessions to build a distributed collaborative visualisation system. Additionally, and to the best of our knowledge, ours is the first attempt to build a collaborative visualisation system for the DH centred around the annotation of uncertainty.

2.4. Citizen Science

Citizen science is a dawning participatory research practice where members of the public and non-professional scientists collaborate with academics to conduct scientific research, usually aided by Information and Communication Technologies (ICT) and other digital tools [27]. These dynamic practices still represent a minority in digital humanities since in general terms, approaches coalescing under the banner of the Arts and Humanities scholarship have a well-established tendency to favour and reward lone scholarship in comparison to life and natural sciences domains [28]. With some recent exceptions, participatory strategies in DH citizen science could benefit from the deeper and more extensive involvement of participants in the effective co-creation of research objects [29], and therefore in dealing with uncertainty in conceptual and epistemological terms.

3. Study Design

In the following sections, we describe two workshops held with DH experts and users, respectively, in which the main task was to identify uncertainties in humanistic digital research objects. Beyond that, we wanted to assess whether the participants could employ the taxonomy presented in Section 2.1 to categorise the uncertainties they would find in the two use-cases we devised for each of these workshops (analysis of literary manuscripts and study of historical recipes, respectively). This was performed with several general objectives in mind: first, we wanted to enable a critical discussion around the idea of uncertainty that could help us understand user needs to drive the design of a hypothetical visual CAS. Second, having a set of predefined rules and agreements is a key component to enable an effective collaboration between multiple peers [30]. Thus, we wanted to assess how workshop participants could establish a consensus around the use of the provided taxonomy to resolve the given exercises. Here, we paid special attention to the identification of situations in which the taxonomy was not expressive enough to cover all the situations that appeared during the exercises and how the participants worked together to resolve these issues. Finally, we aimed to establish a comparison between how the two user groups (experts and users) employed the taxonomy, in an attempt to gather knowledge on how collective reasoning occurs at different levels of user expertise and underlying themes.

4. First Workshop: DH Experts

This section discusses our first approach used in evaluating the uncertainty taxonomy presented in Section 2.1. In this study, we engaged eight humanities researchers (four senior

DH researchers, three PhD students and a professional librarian) in a critical discussion that revolved around the idea of uncertainty in digital research objects, which was in turn enabled by means of two practical exercises centred on literary analysis. The expertise areas of the participants were diverse, ranging from the Creative Arts to Modern History, French Studies, or Cultural Studies. From the conversations that followed the exercise, we could extract an initial set of recommendations that helped us drive the building of a collaborative visualisation platform for the DH and the design of further user studies and evaluations from those presented in Section 6.

4.1. Description

The half-day workshop was hosted by the Trinity College Dublin Centre for Digital Humanities and the consortium member of the project. The workshop was presented as part of the Digital Scholarship and Skills workshop series, an educational initiative of the Centre for Digital Humanities, which aims to introduce participants from a diverse range of backgrounds to digital-research-related skills and tools, with a specific focus on developing a greater understanding and appreciation of how the digital is shaping and influencing scholarship. The workshop was attended by Creative Arts and Humanities researchers and professionals of different levels and disciplines from the Irish Trinity College Dublin and the National University of Ireland Maynooth.

The aims of the workshop were several: firstly, we wanted to introduce the participants to key concepts and terminology related to uncertainty management and modelling, providing them with an overview of our project's aims and objectives. Second, we sought to present the taxonomy of Section 2.1 to the participants and make them use it in a couple of annotation exercises. Finally, we aimed to gather the participants' input to fuel further iterations and refinements of the proposed taxonomy and drive the design process of the digital platform.

4.2. Development

We designed two exercises in which participants had to reason about uncertainty on a specific research object of our choice. The chosen research object was a hard copy of a digitised image of the original writing of *Krapp's Last Tape* by the famous Irish novelist and playwright Samuel Beckett. The file was kindly supplied to us by researchers at the Beckett Digital Manuscript Project (BDMP: <https://www.beckettarchive.org/catalogues/krapp/catalogue>, accessed on 13 October 2021). The document contains many autograph corrections, annotations, and revisions in black and blue inks. This manuscript, labelled by Beckett as 'Typescript II', reflects the changes he made in the first typescript (some of which was revised again) and is a major expansion of that version, corresponding more recognisably to the published text. Introductory material relating to the stage setting has been added, and the verso of page 1 has a long autograph addition which incorporates stage directions concerning A's (Krapp's) movements. Whereas the first typescript was untitled, here, Beckett wrote 'Crapp's Last Tape' at the top of page 1.

4.2.1. First Exercise

In the first exercise, the participants were divided in two groups (Table 1) and were asked to analyse the sample text, identifying uncertainties. We encouraged them to consider what they may be taking for granted in the analysis process and take notes for later discussion. They were instructed to work in whichever way best suited them according to their ground knowledge, for which they were given 25 min. After that, both groups of participants engaged in a collective discussion for 25 min. We observed both the conversation of each group and the final discussion with the participants' permission, taking notes of the most important topics and remarks.

Table 1. Relation of participants and their assigned group.

Participant	Group
Digital Humanities MSc Student	Group 1
Interdisciplinary Research Assistant	Group 1
French Studies Senior Researcher	Group 1
Medieval History PhD Candidate	Group 1
Faculty Librarian	Group 2
History PhD Candidate	Group 2
Creative Arts PhD Candidate	Group 2
Creative Arts Research Fellow	Group 2

4.2.2. Second Exercise

In the second exercise, the participants were introduced to the uncertainty taxonomy of Section 2.1. Again, they were placed in the same groups of exercise 1, and asked to revisit the annotations of uncertainty they had previously made on the text, applying the taxonomy where appropriate. They were encouraged to note where the taxonomy was unclear and similarly, to note those examples of uncertainty that they felt did not fit the taxonomy headings and descriptions. Again, a group discussion followed the exercise and they were asked to fill in a post-task questionnaire that gathered their impressions about the workshop and the taxonomy in a more formal manner. The questionnaire was anonymous and was filled in online by the participants after the event (see Supplemental Materials). Their answers are briefly discussed in the next section.

4.3. Main Outcomes

4.3.1. General Remarks

In general, the participants found the workshop to be relevant to their research activity (4.25 ± 0.66 on a 5-point Likert scale) and their self-reported general level of satisfaction with the experience was high (4.5 ± 0.70). Generally, we obtained positive comments related to the delivery of the workshop and the usefulness of the exercises (6/8), with some participants stating their interest in more follow-up activities on the topic. In this regard, one participant wrote that “it was both fascinating and interesting to learn more about the new research in this area”, while another one wrote that “the hands-on aspect of the workshop was the most beneficial”. Although it was not used in the workshop, two participants mentioned that they would have liked to know more about the technical implementation of the workshop’s ideas on a platform, if given more time.

4.3.2. Uncertainty in Humanities Research

When asked specifically about how the main topics of the workshop could be applied to their daily research activities, we found the participants to be concerned about the role of uncertainty in the analysis of primary sources. In this regard, one participant expressed that “transcriptions from original manuscripts can be modified and printed not in verbatim, which is when the uncertainty matters in my study. So the workshop encouraged me to think of layers of edition process”, an appreciation that is also discussed in recent work by Lamqaddam et al. [31]. Another participant, whose research focused on the use of web archives for digital scholarship, pointed to an important void in the literature of her field regarding uncertainty management. The participant mentioned several challenges and use-cases in which the adoption of appropriate decision-making support techniques would be beneficial for her. Concretely, this researcher mentioned problems related to the effective assessment of completeness in sets of different website snapshots collected by different actors, whose “large size and (sometimes) erratic temporal distribution represent major hurdles”. Next, she highlighted the “lack of metadata, finding aids, and information on the scope and parameters employed to generate a given snapshot”, which would help her answer important questions as “what was collected and why, or what was left out

and why". Finally, she went back to the matter of increasing data volumes, for which she identified a lack of training in humanities researchers, which is inclusive of "ethical paradigms to approach data with uncertainties".

4.3.3. Visual Analytics and Decision-Making in the Humanities

Regarding the opportunities that visual analytics focusing on uncertainty representation could offer in their particular research contexts, several participants made explicit mention of the potential of visual analytics tools as teaching and communicative instruments to convey their work. For example, one participant expressed that "any opportunity for historians to 'show their work' when dealing with sources not only demystifies what we do to the public but it also invites others into what we do. I absolutely love the idea of implementing a chart allowing us to 'show' how and why we make decisions", while another one pointed out the usefulness of visualisation "to better document the research process and to appeal to different audiences". In a different aspect, another participant saw potential applications of this kind of tool in "isolating divergent or distinctive documents in large datasets of unstructured data (e.g., text) for close reading".

5. Lessons Learned

After a careful revision of the materials and conversations generated during the workshop, we could extract a series of recommendations for digital humanists developing digital research platforms that deal with uncertainty in the data analysis pipeline. These were classified as follows:

5.1. Uncertainty Modelling

An important issue that was raised during the conversations was that working with multi-level, non-binary error scales (i.e., "certain", "rather incorrect") as opposed to employing binary ones (i.e., "credible" vs. "non-credible", "precise" vs. "imprecise") may be beneficial to enable an appropriate assessment and communication of self-reported judgements made by multiple actors on a given digital research object. Particularly, this was revealed when participants were shown the annotations made by the other group, highlighting instances of conflicting annotations, which made some of them modulate their initial judgements. According to the input received from the participants, it would be interesting that the hypothetical collaboration system allowed them to indicate which annotations, according to their own judgement, are more or less "uncertain" (e.g., "correct") without obscuring the other alternatives. However, the extent to which this way of collaboration may be adopted by the research community in the future was called into question, as the complexity and difficult usage of such a system could greatly hinder the potential benefits. Thus, extensive user validation will be required to corroborate that a hypothetical system with these characteristics can indeed accelerate collaborative data analysis in the DH. Finally, it was also agreed that the taxonomy needs to provide for the identification of both the type of uncertainty and the source of uncertainty in the research process. To that end, a matrix-like taxonomy may be of the most use to researchers in this context. An example uncertainty matrix resulting from our conversations with the participants is presented in Figure 2.

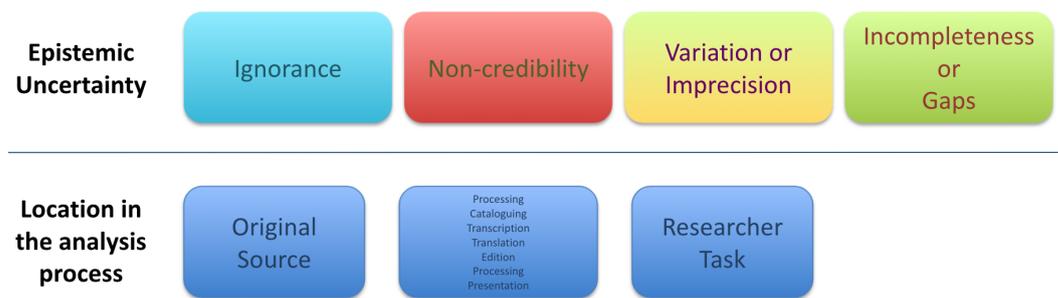


Figure 2. The two dimensions of the matrix of uncertainty for the DH that was built during the first workshop: type of epistemic uncertainty (**top**, as per Simon et al.’s types) and location in the DH data analysis process where it originates (**bottom**).

5.2. Design of DH Platforms

Regarding the design of DH platforms, it was made clear during the discussion that the teams behind them should adopt strategies to make the underlying algorithms more understandable, making the uncertainty introduced by them in the system more evident and noticeable, rather than hiding it. Some participants suggested the idea that these strategies should go far beyond the mere production of documentation and tutorials, but rather the implementation of methods specifically designed to increase algorithmic literacy in the user. When presented with potential solutions, they all agreed that the use of interaction and visual analytics techniques specifically designed to open the algorithmic black-box [32,33] would be extremely helpful in a DH research context. For this, they stressed the importance of providing visualisations that are able to link the annotation task of individual documents to the distant reading of the collection base that results from it, an observation that is in line with previous works in the field [15].

Finally, an issue that was repeatedly raised during the discussions was the authority of the annotator. It was recommended that the platform’s users should be linked to a public profile (e.g., ORCID). Beyond that, users should be able to set a personalised level of confidence on annotations made by another peer. In the case of academic annotators, profiles could be linked to their ORCID. Use-cases should be provided to researchers to demonstrate how the platform enhances research. This is particularly true of the historical research process, where it was felt that the platform may not be able to make a meaningful contribution.

5.3. Social Considerations

Finally, participants also stressed the value of acknowledging that a number of social barriers exist that may greatly impact the outcomes of our proposal. For example, the issue of data citation was a main concern to the participants because, as some of them argued, “humanistic data curation and preparation is both a time-consuming and epistemically-demanding process”. A digital system using such data naturally raises issues of both credit and bias in the authors making use of such systems, thus requiring robust mechanisms for traceability, attributability, and accountability to be built-in into it. Furthermore, it was also mentioned that there is a general tendency among arts and humanities scholars to favour and reward lone scholarship. To overcome this issue, the participants repeated the necessity to build digital systems that encourage humanities researchers to share their data for others to work with, rewarding and recognising the labour involved in the creation of original datasets. For this, the adoption of mechanisms for automatic versioning and authorship attribution of corpora was regarded as highly relevant.

6. Second Workshop: DH Users

After the initial expert evaluations, and based on our internal reflections on the results, we decided to conduct two quantitative and qualitative user studies centred around the annotation task of historical documents. The goal of these experiments was twofold: firstly,

we wanted to obtain insight on how users employed our proposed uncertainty taxonomy in a real-world scenario using data they were familiar with. Second, we wanted to inform the development of a digital annotation tool focused on uncertainty quantification, which is one of the goals of our project. In this regard, we prepared a simple annotation prototype (Figure 3, left) that was employed in certain controlled tasks during the interactions to obtain first-hand early feedback from the participants. However, and given the experimental state of the prototype at the time, we chose not to base the workshops on the digital prototype, which was moved to a second plane. Instead, and after several internal discussions, we decided to conduct the experiments at a lower fidelity level using pen and paper and make participants use the digital tool just to transcribe the paper-based annotations to the digital medium using a coaching method inspired by Mack and Robinson [34]. With this approach, we wanted to obtain early feedback on the prototype while avoiding potential interaction problems that could have appeared due to the experimental nature of the prototype and the expected low level of computational literacy of our users.

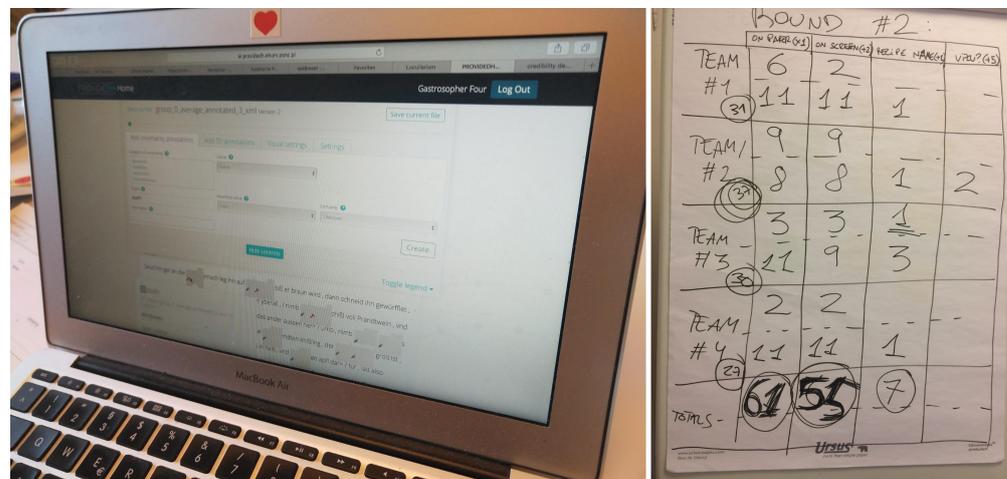


Figure 3. (Left) Annotation prototype shown running on the laptop of one of the participants. The prototype, that was used by the users under controlled conditions in the second part of the workshops, offered a simple interface to attach annotations to TEI documents employing the proposed uncertainty taxonomy. (Right) An example of an annotation identified as “other/unknown” in the recipe transcriptions.

6.1. Description

In order to recruit participants for our study, we collaborated with stakeholders from the citizen science community working on the historical recipe collection “Cooking up Salzburg” (“Salzburg zu Tisch”), a citizen-driven initiative led by the Centre for Gastrosophie at the History Department of the University of Salzburg [35]. The *Gastrosophie* (<https://www.gastrosophie.at/>, accessed on 13 October 2021) community is settled around a scientific project dealing with medieval recipes from the Salzburg region. The community is involved in gaining and enriching data (and gets trained for this task) but also in experimentation by re-cooking the recipes together with experts. The *Gastrosophie* group is therefore strongly connected to a scientific environment. It is a rather small group but the members seem to be quite diverse in terms of gender and also age.

The project manages a compelling collection of circa 8700 handwritten originals describing historical recipes from the area of Austria written in non-standard German (Austrian–German). In order to leverage the interactions with our target users, we relied on several participatory design (PD) methods [36] such as gamification (Figure 3, right) to drive the interventions, which we had successfully employed in the past in other DH-related initiatives involving lay users [37].

6.2. Development

In total, we held two workshops (WS1 and WS2) with volunteers from the *Gastrosophie* project. Both workshops had an approximate duration of 120 min. The total number of distinct participants was 20 (WS1: 10 female, 2 male; WS2: 6 female, 2 male). Both workshops were organised similarly: firstly, the facilitators provided an introduction to the PROVID-EDH project and uncertainty taxonomy, which was led by the activity of self-accreditation by roles and skills in the context of DH research. According to the self-reported levels of expertise and skills, the participants were distributed into four groups of two to four people with similar expertise levels. Each group was handed several printed recipes (more than they could possibly complete in the allocated time of 20 min), and instructed to read the recipes and annotate, using a paper template (Figure 4) and coloured stickers and pens, the passages that they thought could be linked to each of the four types of uncertainty described in the taxonomy. During the workshops, and especially during the annotation exercises, we performed extensive note-taking to capture the participants' activity. In this first part, we followed a think-out-loud (TOL) protocol [38], which is a common practice in UI and UX design, and also in DH [39].

The figure shows a paper template for annotation. At the top, there are two input fields: "Team ID: _ _ _ _ _" and "Recipe number: [_]". Below these are three main sections. The first section is "Category of uncertainty" with four checkboxes: "Ignorance" (blue), "Credibility" (red), "Imprecision" (yellow), and "Incompletedness" (green). The second section is "Locus" with a "Name" input field. The third section is "Certainty" with five checkboxes: "Very high", "High", "Medium", "Low", and "Very Low". Below these are three input fields: "Target", "Asserted value", and "Certainty". A "Description" field is located below the "Target" field. At the bottom, there is a legend with 15 checkboxes and icons representing different entities: Date, Object, PlaceName, ProductionMethod, Event, ObjectName, Country, Location, Org, Time, Geolocation, Person, Ingredient, Occupation, and Utensil. Dotted arrows indicate that the "Asserted value" and "Certainty" fields are linked to the legend.

Figure 4. Paper template provided to participants in the second workshop.

Finally, for each annotation, they were told to add a small descriptive piece of text explaining the reason why they had annotated a given set of words (Figure 5).

The annotation task was followed by a group discussion in which the participants reflected on the most prominent uncertainties found. Then, the annotation prototype was presented to the participants, explaining its functionality through a couple of examples. After that, the participants were asked to translate the annotations made in the first part of the workshop to the digital medium using the annotation prototype using their own laptops. Furthermore, the participants were given detailed instructions on how to access the tool, log in with the predefined users and passwords, and access the demo project. Finally, we held another informal group discussion with them, focusing on the aspects of the prototype they considered to be more useful or that needed to be redesigned.

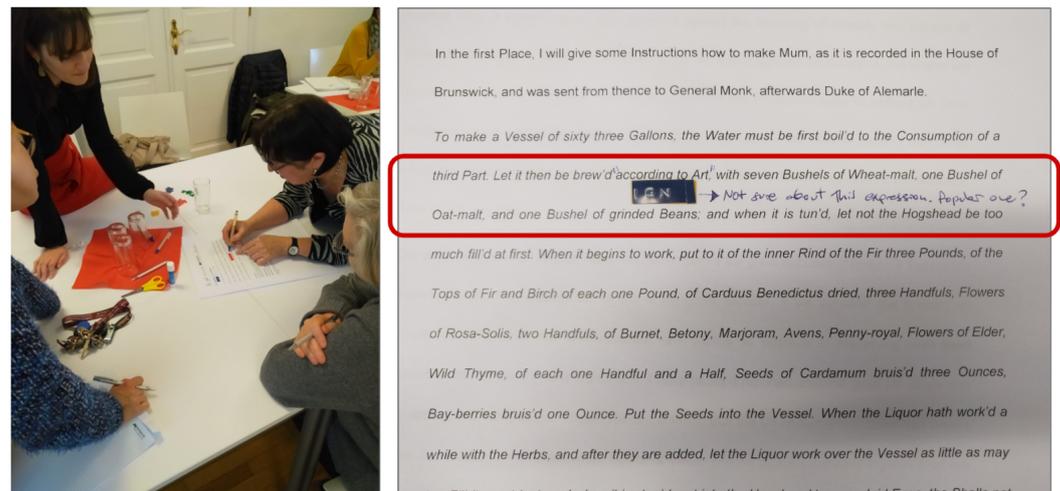


Figure 5. (Left) Participants engaged in the process of annotation. (Right) An example of an uncertainty annotation identified as “other/unknown” in the recipe transcriptions.

6.3. Results

In order to better understand how participants made use of the uncertainty taxonomy during the programmed exercises, we performed a mixed-methods analysis of their annotations, attending to four distinct dimensions: (1) language employed in the annotations, (2) usage of uncertainty categories, (3) length of annotations and annotated portions of text, and (4) self-reported confidence levels. In the first case, our aim was to measure to what extent the participants were able to adopt the uncertainty taxonomy for resolving the exercise, and how they employed its four categories of uncertainty, enabling us to know more about its potential deficiencies and shortcomings. In the second case, we wanted to know if the two groups employed the categories in the taxonomy in the same manner, that is, the categories were used to mark similar cases of uncertainty in different parts of the dataset by different people. With this, we wanted to verify whether the use of a taxonomy might be effective in establishing a reliable communication channel between collaborators of the same project. In the third case, we sought to gain knowledge on the length of the portions of the text that were annotated, which could inform the design of visual elements for the distant and close reading of uncertainty in text documents. Finally, and following the recommendations of Section 5.1 on the use of multi-level error scales to express levels of subjective confidence on uncertainty judgements, in the second workshop, we asked participants to add those to their annotations using a 5-point Likert scale ranging from 1, “Very low uncertainty” (“I am very confident about my statement”) to 5, “Very high uncertainty” (“I am not confident at all”). With this, we wanted gain insight on potential hidden connections between uncertainty categories and the self-reported confidence levels employed in the same annotations.

6.3.1. Categories of Uncertainty Annotations

In total, we obtained 118 uncertainty annotations from 15 recipes (WS1: 50, WS2: 60; avg: 7.87 annotations per recipe, recipe length: 100–200 words). A distribution per category is shown in Figure 6. Looking at the figure, it can be seen that the “ignorance” category, which was related to a perceived lack of subjective knowledge regarding words appearing in the texts, was the most used by a large margin in both interventions (43% and 66% of the total number annotations, respectively).

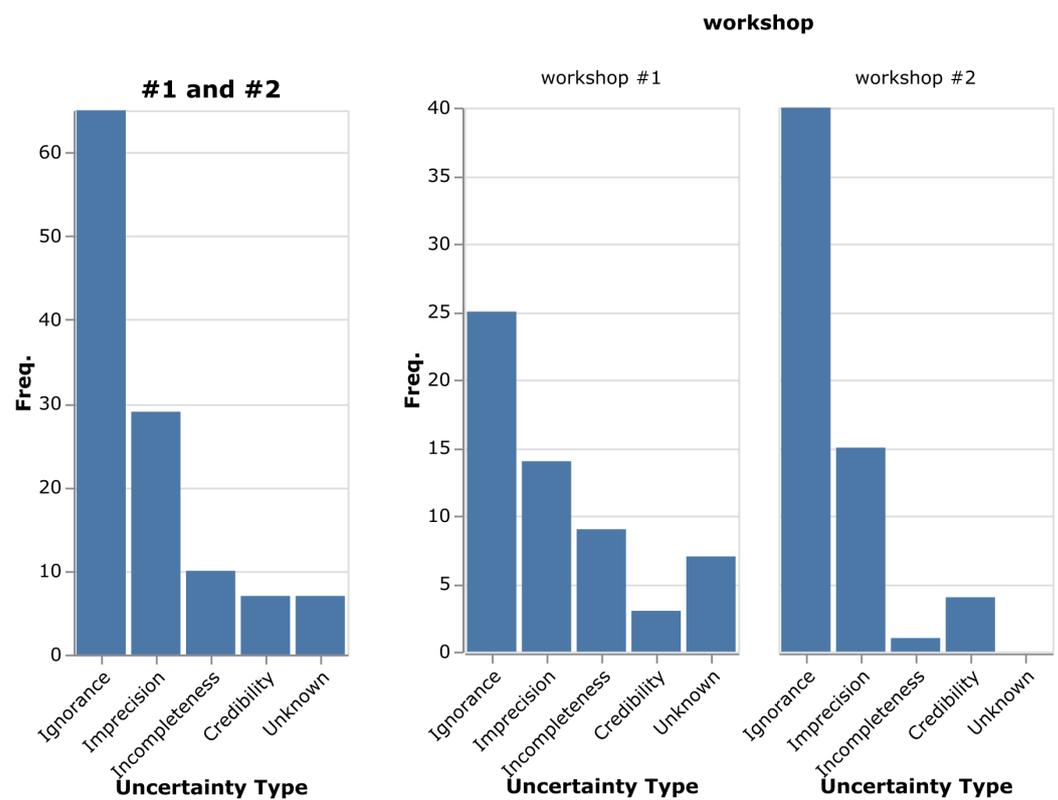


Figure 6. (Left) Distribution of uncertainty annotations by category. (Right) Breakdown by workshop. By a large margin, the “ignorance” category was the most used by participants in our study, followed by “imprecision”. In addition, participants made a total of 7 annotations that they could not link to any of the categories in the taxonomy.

A similar effect occurred with the category “imprecision” (in this case, related to unclear or general indications on recipes, mainly), which appeared as the second most used in both workshops. The other two categories, “incompleteness” (around 8.5%) and “credibility” (around 0.5%), were the least employed by the participants, indicating a clear bias towards usage of the first two. As we see later, we relate this skew to two main factors: the nature of the task and the user’s level of expertise. Interestingly, in some situations, the users were unable to choose a specific category. In these cases, we instructed them to leave it as “unknown”. This finding is discussed in the next section.

6.3.2. Category Usage

As previously introduced, we wanted to gain a better understanding of how participants employed the categories in the proposed taxonomy to mark uncertainty in the recipe texts, for which we relied on (1) the comments they made during the annotation session and (2) the actual text of the annotation they provided. A sample of common expressions employed in the annotations is presented in Table 2. The full set of annotations can be found in the Supplemental Materials of this paper.

Table 2. Example annotation text for each uncertainty category (translated into English from German).

Category	Example Annotations
Ignorance	<ul style="list-style-type: none"> • “pot or broth ingredient?” • “unknown cooking utensil” • “maybe blue?”
Imprecision	<ul style="list-style-type: none"> • “unclear quantity” • “maybe transcription error, maybe means butter” • “wrong transcription; it should be: salze es, meaning to add salt”
Incompleteness	<ul style="list-style-type: none"> • “vague description of preparation” • “missing ingredient + missing mode of preparation” • “what? missing word”
Credibility	<ul style="list-style-type: none"> • “pigs feet in a cake?” • “the application is difficult; it’s more a set of guidelines than a recipe” • “contradiction in the instruction”
Unknown	<ul style="list-style-type: none"> • “prior knowledge necessary → needs new category: implicit knowledge; knowledge that contemporary cooks need to have.” • “title not from original source” • “ingredient not needed (?)”

7. Discussion

In this section, we discuss our main findings regarding how the participants of our workshops employed the taxonomy to resolve the given use-cases, along with other observations that we considered worthy of discussion here.

7.1. On the Use of Categories

Ignorance: In a first inspection, we learned that the majority of annotations labelled under the “ignorance” category referred to non-standard terms appearing in the original texts for which participants had trouble grasping their meaning. Sometimes, the participants tried to guess a term’s meaning from its context, reflecting it in the annotation (e.g., “maybe blue?”). In a couple of cases, the users’ guess suggested the misunderstanding could be linked to transcription errors, which could be more aligned to a case of “imprecision”, as we explain next. In any case, the general consensus was the use of “ignorance” to report a self-perceived lack of information to decode a term’s meaning, and annotations of these kind were mainly employed at the word level.

Imprecision: Regarding annotations labelled as “imprecision”, they mentioned issues with vagueness in numerical values (e.g., imprecise quantities of an ingredient or cooking times), which made it hard for them to reproduce the recipe. This also occurred when a given term was considered too general to follow the recipe in a correct manner. In addition, several participants also showed incredulity when encountering certain “unexpected” terms or expressions, which they sometimes attributed to an improper use of language or even to transcription errors, as had occurred in the previous case. Although we could not experimentally verify under what exact conditions participants classified these cases as “imprecision” or “ignorance”, in the conversations that followed the experiment, we got the impression that this decision was related to the user’s level of confidence in her statement. In addition, we also observed a certain level of conceptual overlap between “imprecision” and the other two categories (“incompleteness” or “credibility”) of annotations in cases when they felt one or more cooking steps was given in an incorrect manner, or even when

they identified some missing components in the recipe, which was the most common use-case for the “incompleteness” uncertainty category. Another important remark is that although imprecision could be observed by participants in both the Beckett and recipe annotation exercises, each group *interpreted* it in the manner that could be best applied to the problem and data they were dealing with. For example, in the Beckett exercise, autographs and corrections where the original content could not be clearly seen were marked as imprecision. However, in the recipe annotation exercises, imprecision was associated more often with unclear statements about amounts (of time or ingredients). Thus, there is a clear link between an activity’s main aims and tasks and the interpretation the participants performing these tasks will make of a taxonomy, an effect that was also observed in the other categories.

Incompleteness: Annotations labelled as “incompleteness” described situations wherein the recipe was missing important ingredients or procedures that the participants considered absolutely necessary to reproduce the recipe, or even when two steps were given in an incorrect order. In the majority of these occasions, participants were unable to link their annotation to an specific part of the recipe’s text, and they mentioned they preferred to associate it to the whole object (the recipe itself). This was a use-case that we had not contemplated in the preparation of the workshops, and pointed us to an important matter related to the actual annotation lengths, which it is further discussed in Section 8.

Credibility: The “credibility” category (which was the least used of all) depicted annotation scenarios in which the participants felt that despite the fact that they could understand the text correctly, it presented logical flaws and contradictions in the description of the recipe that make it very hard or impossible to reproduce. Beyond that, in certain cases, the users could link this uncertainty type to unexpected appearances of terms that seemed to be “out of place”, (e.g., swearwords or very unusual ingredients) which cast doubts on the veracity of the whole recipe. Additionally, in some cases, the participants followed an hermeneutical approach [40] to resolve the issue, producing new “correct” versions of the recipe according to their own beliefs. Another interesting discovery was that users had reversed the scale in the “credibility” annotations, choosing higher values when they were *more certain* about their statement. In conversations with the participants, they told us that this naming was confusing, and it should be changed to “non-credibility” to equate its sense with the other category names, a decision that we will adopt in future studies.

Unknown: As was introduced in the previous sections, in some cases, the participants were unable to decide what category they should assign to a particular uncertainty found in the recipe’s text. The cases mainly referred to dialect-related issues in which the text employed non-standard German expressions or words. When asked about why they could not link these cases to any of the existing categories, the participants acknowledged that whereas they could have employed “imprecision” (i.e., as in linguistic imprecision), they felt the conceptual difference with other items in the category was high enough to put them in a separate category. In some other cases, the reason was that the recipe mentioned *too many* (or unneeded, according to their understanding) ingredients. However, in the talks that followed the exercise, they admitted that they could have placed these cases as manifestations of “credibility”, since encountering these unexpected terms made them question the integrity of the recipe as a whole.

7.2. Self-Reported Confidence Levels

Although the data are scarce, the discussions we held with the participants led us to think that there might be a connection between how often an uncertainty type can be detected by a group of users (the “surprise factor”) and their degree of self-reported confidence, which in turn could also be connected to how well the use-case can be fitted into the category (or vice versa). Undoubtedly, the kind of credibility annotations that were made in these examples were more prone to induce higher confidence values in the users (a contradiction can be detected or not). Similarly, to state that something is “incomplete”, one must have a notion of what the whole is, or otherwise such an observation could be simply

classified as “ignorance” or “imprecision”, or even missed altogether. The classification in this case also depends on one of two possible outcomes: either the user is aware of the whole or they are not. This situation makes a good case for collaboration, since another user working on the same data could be aware of this difference (e.g., because they have more expertise on the matter), and thus could attach their perspective to the previous annotation made by the first user, changing its type, a term that we have determined to name as “assertion”.

7.3. On the Use of a Taxonomy

In light of our results, the adoption of an uncertainty taxonomy seems to be a key element to enable effective collaboration in digital humanities platforms, because we argue that it represents an internal agreement between the participants on how to interpret the categories to derive a classification of the uncertainty found in a digital research object in a given context. Thus, the establishment of this consensus needs to be exploited by the system to enable an interpersonal communication channel through which the stakeholders’ different perspectives and judgements can be fixed into the data and transmitted over time and space. In this regard, different stakeholders of a research project collaborating on a digital research object will be responsible for establishing a consensus uncertainty taxonomy at the start of the project, which will be closely linked to their level of expertise, and to the aims and tasks of the project.

7.4. Dynamic and Per-Project Taxonomies

According to our observations, rather than being a static process that occurs only at the initial stages of a project, the arrival at a consensus uncertainty taxonomy is a dynamic one that extends in time as more use-cases of the taxonomy are covered and more users join the discussion around a research object. Thus, an effective system will accelerate the refinement of this consensus taxonomy during the entire lifetime of a project. This could be derived from the observations made in cases when participants have differing opinions on what category should be assigned to a certain passage of text. In these situations, they often had to stop progressing through the exercise and have a short debate on how to apply the taxonomy to the given use-case. Once the discrepancies were resolved and an internal consensus in the group could be established, they were able to continue with the activity, applying this agreement in the next similar situations they would face. Furthermore, it is worth noting that potential refinements of a taxonomy are not limited to the addition of more categories, but also to the renaming of existing ones. This was observed, for example, in the workshop of Section 4, in which the participants decided to rename “credibility” to “non-credibility”, to make it compatible with the other categories which were given in a negative form (“**im**precision”, “**in**completeness”), a change that made more sense to them as a group.

8. Conclusions and Future Work

During the course of this research, we identified several limitations that we aim to overcome in future work. These are discussed hereafter.

One main shortcoming of our research is that it only focuses on the collection of uncertainty during the annotation task. However, one important outcome of our project is to measure to what extent the inclusion of this uncertainty may leverage DH scholars’ collaborative work, i.e., *does the inclusion of uncertainty lead them to work more efficiently?* Since the goals of humanistic research are fundamentally different from those of experimental sciences, it seems important to establish clear metrics that can measure this hypothetical increase in efficiency or, perhaps more importantly, in the quality of the results. For example, it is widely assumed that humanities scholars often seek to produce critical analyses that may inspire others to do the same [31]. Accordingly, a good collaborative system in this context would be one that allows a wider range of different perspectives to be fixed to the data in an integrative approach, in turn motivating critical analyses that take more of them

into account. To this end, we plan on adopting and adapting existing specific evaluation techniques that are already available in the visualisation literature [41,42].

During the sessions, the problems of overlapping and clogging were brought up by the participants. For example, in our preliminary prototype, we employed a simple visual coding scheme for uncertainty annotations that relied on the use of coloured tags. Although this method is rather straightforward and has many advantages, such as enabling users to make correct estimations of the amount of uncertainty present in the text [43], it also presents certain scalability issues that render it inadequate for a multi-user collaborative environment. This is due to the observation that different uncertainty categories are typically associated with distinct text lengths. For example, in the recipe annotation exercise, we could see that annotations of “imprecision” were mostly linked with single words, whereas “non-credibility” annotations referred to larger pieces of text. This creates a good opportunity to allow users to dynamically filter the type or types of uncertainties (as per their established taxonomy) they want to see at a given time. The system could then employ distinct visualisation techniques to display uncertainty annotations that would rely on these differences in length to display them in the best possible manner. In this regard, and although the proposed uncertainty annotation tasks were centred on single documents, we plan to include distant reading approaches to facilitate the navigation and filtering of documents based on the nature of the annotations they contain.

Another important shortcoming is that our study only considers epistemic uncertainty performed by human annotators. However, as the use of ML, NLP, and text-mining techniques has become widespread in DH, there exists a good opportunity to produce machine-annotated editions of large corpora that would be impossible to obtain via manual means. As many of these algorithms are probabilistic in nature (e.g., topic models), this creates a good opportunity to introduce their results as aleatoric uncertainty in the system. Under this approach, the users of the system could make new annotations (or assertions) derived from the results of these algorithms and those from other peers, effectively creating a sort of Bayesian belief network [44] that could be exploited to produce more advanced types of analyses.

In this paper, we presented two participative design user studies aiming to evaluate a taxonomy of textual uncertainty for use in supporting collaboration in a DH research environment. Through these experiences, we were able to gain interesting knowledge on how this taxonomy can enable collective reasoning between users at different levels of expertise and research themes. Beyond that, we could obtain first-hand feedback on a simple prototype that employed this taxonomy to drive the annotation task in a collaborative manner. In an attempt to fill the current need to bring uncertainty to the surface of the DH research process, we extracted a set of observations that we are actively employing to enhance our digital platform (currently available at <https://providedh.ehum.psnr.pl/>, accessed on 13 October 2021), and that we shared here with the rest of the community.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/info12110436/s1>.

Author Contributions: Conceptualization, A.B.-S., M.D., E.W.-V., J.E. and R.T.; methodology, A.B.-S., M.D., E.W.-V., J.E. and R.T.; software, A.B.-S. and R.T.; validation, E.W.-V., J.E. and R.T.; formal analysis, A.B.-S.; investigation, A.B.-S. and J.E.; resources, E.W.-V., J.E. and R.T.; data curation, A.B.-S., A.R. and M.D.; writing—original draft preparation, A.B.-S. and A.R.; writing—review and editing, A.B.-S., A.R., M.D., E.W.-V., J.E. and R.T.; visualization, A.B.-S.; supervision, E.W.-V., J.E. and R.T.; project administration, R.T.; funding acquisition, E.W.-V., J.E. and R.T. All authors have read and agreed to the published version of the manuscript.

Funding: This publication is part of the i+D+i project PCIN-2017-064, funded by MCIN/AEI/10.13039/501100011033/and co-funded by the European Union. It was also funded by the Irish Research Council under grant CHISTERA, and the Austrian Science Fund (FWF) under grant number I 3441-N33.

Acknowledgments: The authors want to thank the Irish Trinity College Dublin, the National University of Ireland Maynooth, the Centre for Gastrosophy at the History Department of the University of Salzburg, Marlene Ernst, Amelie Dorn, Renato Souza, and Enric Senabre-Hidalgo and the Open Access Funding by the Austrian Science Fund (FWF) for their collaboration in this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Drucker, J. Humanities Approaches to Graphical Display. *Digit. Humanit. Q.* **2011**, *5*, 1–21.
- Windhager, F.; Salisu, S.; Mayr, E. Exhibiting Uncertainty: Visualizing Data Quality Indicators for Cultural Collections. *Informatics* **2019**, *6*, 29. [[CrossRef](#)]
- Benito-Santos, A.; Sánchez, R.T. A Data-Driven Introduction to Authors, Readings, and Techniques in Visualization for the Digital Humanities. *IEEE Comput. Graph. Appl.* **2020**, *40*, 45–57. [[CrossRef](#)]
- Martin-Rodilla, P.; Gonzalez-Perez, C. Conceptualization and Non-Relational Implementation of Ontological and Epistemic Vagueness of Information in Digital Humanities. *Informatics* **2019**, *6*, 20. [[CrossRef](#)]
- Edmond, J. Strategies and Recommendations for the Management of Uncertainty in Research Tools and Environments for Digital History. *Informatics* **2019**, *6*, 36. [[CrossRef](#)]
- Chen, C.M.; Tsay, M.Y. Applications of Collaborative Annotation System in Digital Curation, Crowdsourcing, and Digital Humanities. *Electron. Libr.* **2017**, *35*, 1122–1140. [[CrossRef](#)]
- Stewart, M.; Liu, W.; Cardell-Oliver, R. Redcoat: A Collaborative Annotation Tool for Hierarchical Entity Typing. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations, Hong Kong, China, 3–7 November 2019; pp. 193–198. [[CrossRef](#)]
- Steiner, C.M.; Agosti, M.; Sweetnam, M.S.; Hillemann, E.C.; Orio, N.; Ponchia, C.; Hampson, C.; Munnely, G.; Nussbaumer, A.; Albert, D.; et al. Evaluating a Digital Humanities Research Environment: The CULTURA Approach. *Int. J. Digit. Libr.* **2014**, *15*, 53–70. [[CrossRef](#)]
- Gius, E.; Meister, J.C.; Meister, M.; Petris, M.; Bruck, C.; Jacke, J.; Schuhmacher, M.; Gerstorfer, D.; Flüh, M.; Horstmann, J. CATMA. Zenodo. 2021. Available online: <https://zenodo.org/record/5015305> (accessed on 13 October 2021). [[CrossRef](#)]
- Bradley, A.J.; El-Assady, M.; Coles, K.; Alexander, E.; Chen, M.; Collins, C.; Jänicke, S.; Wrisley, D.J. Visualization and the Digital Humanities: Moving Toward Stronger Collaborations. *IEEE Comput. Graph. Appl.* **2018**, *38*, 26–38. [[CrossRef](#)] [[PubMed](#)]
- Jänicke, S. Valuable Research for Visualization and Digital Humanities: A Balancing Act. In Proceedings of the 1st Workshop on Visualization for the Digital Humanities (VIS4DH), Baltimore, MD, USA, 24 October 2016.
- Hullman, J.; Qiao, X.; Correll, M.; Kale, A.; Kay, M. In Pursuit of Error: A Survey of Uncertainty Visualization Evaluation. *IEEE Trans. Vis. Comput. Graph.* **2018**, *25*, 903–913. [[CrossRef](#)] [[PubMed](#)]
- Hullman, J. Why Authors Don't Visualize Uncertainty. *IEEE Trans. Vis. Comput. Graph.* **2020**, *26*, 130–139. 2934287. [[CrossRef](#)]
- Isenberg, P.; Elmqvist, N.; Scholtz, J.; Cernea, D.; Ma, K.L.; Hagen, H. Collaborative Visualization: Definition, Challenges, and Research Agenda. *Inf. Vis.* **2011**, *10*, 310–326. [[CrossRef](#)]
- Cheema, M.F.; Jänicke, S.; Scheuermann, G. AnnotateVis: Combining Traditional Close Reading with Visual Text Analysis. In Proceedings of the 1st Workshop on Visualization for the Digital Humanities (VIS4DH), Baltimore, MD, USA, 24 October 2016.
- Therón, R.; Benito, A.; Santamaría, R.; Losada, A.G. Toward Supporting Decision-Making Under Uncertainty in Digital Humanities with Progressive Visualization. In Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality, Salamanca, Spain, 24–26 October 2018; ACM: New York, NY, USA, 2018; pp. 826–832. [[CrossRef](#)]
- Therón Sánchez, R.; Benito-Santos, A.; Santamaría Vicente, R.S.; Losada Gómez, A. Towards an Uncertainty-Aware Visualization in the Digital Humanities. *Informatics* **2019**, *6*, 31. [[CrossRef](#)]
- Fisher, P.F. Models of Uncertainty in Spatial Data. *Geogr. Inf. Syst.* **1999**, *1*, 191–205.
- Smithson, M. *Ignorance and Uncertainty: Emerging Paradigms*; Cognitive Science; Springer: New York, NY, USA, 1989.
- Simon, C.; Weber, P.; Sallak, M. *Data Uncertainty and Important Measures*; John Wiley & Sons: Hoboken, NJ, USA, 2018.
- Joshi, A.; Kale, S.; Chandel, S.; Pal, D.K. Likert Scale: Explored and Explained. *Br. J. Appl. Sci. Technol.* **2015**, *7*, 396. [[CrossRef](#)]
- Boone, H.N.; Boone, D.A. Analyzing Likert Data. *J. Ext.* **2012**, *50*, 1–5.
- Kozak, M.; Rodriguez, A.; Benito-Santos, A.; Therón, R.; Doran, M.; Dorn, A.; Edmond, J.; Wandl-Vogt, E. Analyzing and Visualizing Uncertain Knowledge: The Use of TEI Annotations in the PROVIDEDH Open Science Platform. *J. Text Encoding Initiat.* **2021**, in press.
- Binder, F.; Entrup, B.; Schiller, I.; Lobin, H. Uncertain about Uncertainty: Different Ways of Processing Fuzziness in Digital Humanities Data. In Proceedings of the Digital Humanities, Lausanne, Switzerland, 8–12 July 2014; Volume 15.
- Mwalongo, F.; Krone, M.; Reina, G.; Ertl, T. State-of-the-Art Report in Web-Based Visualization. *Comput. Graph. Forum* **2016**, *35*, 553–575. [[CrossRef](#)]
- Badam, S.K.; Elmqvist, N. PolyChrome: A Cross-Device Framework for Collaborative Web Visualization. In Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces, ITS '14: Interactive Tabletops and Surfaces, Dresden, Germany, 16–19 November 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 109–118. [[CrossRef](#)]

27. Bonney, R.; Cooper, C.B.; Dickinson, J.; Kelling, S.; Phillips, T.; Rosenberg, K.V.; Shirk, J. Citizen science: A developing tool for expanding science knowledge and scientific literacy. *BioScience* **2009**, *59*, 977–984. [[CrossRef](#)]
28. Hecker, S.; Garbe, L.; Bonn, A. *The European Citizen Science Landscape—A Snapshot*; UCL Press: London, UK, 2018.
29. Carletti, L.; Coughlan, T.; Christensen, J.; Gerber, E.; Giannachi, G.; Schutt, S.; Sinker, R.; Denner dos Santos, C. Structures for knowledge co-creation between organisations and the public: (cop2014). In Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, Baltimore, MD, USA, 15–19 February 2014; pp. 309–312.
30. Conoley, J.C.; Conoley, C.W. Why Does Collaboration Work? Linking Positive Psychology and Collaboration. *J. Educ. Psychol. Consult.* **2010**, *20*, 75–82. [[CrossRef](#)]
31. Lamqaddam, H.; Moere, A.V.; Abeele, V.V.; Brosens, K.; Verbert, K. Introducing Layers of Meaning (LoM): A Framework to Reduce Semantic Distance of Visualization In Humanistic Research. *IEEE Trans. Vis. Comput. Graph.* **2021**, *27*, 1084–1094. [[CrossRef](#)]
32. Mühlbacher, T.; Piringer, H.; Gratzl, S.; Sedlmair, M.; Streit, M. Opening the Black Box: Strategies for Increased User Involvement in Existing Algorithm Implementations. *IEEE Trans. Vis. Comput. Graph.* **2014**, *20*, 1643–1652. [[CrossRef](#)] [[PubMed](#)]
33. Stolper, C.D.; Perer, A.; Gotz, D. Progressive Visual Analytics: User-Driven Visual Exploration of In-Progress Analytics. *IEEE Trans. Vis. Comput. Graph.* **2014**, *20*, 1653–1662. [[CrossRef](#)]
34. Mack, R.; Robinson, J.B. When Novices Elicit Knowledge: Question Asking in Designing, Evaluating, and Learning to Use Software. In *The Psychology of Expertise: Cognitive Research and Empirical AI*; Hoffman, R.R., Ed.; Springer: New York, NY, USA, 1992; pp. 245–268. [15](#). [[CrossRef](#)]
35. Ernst, M. Salzburg zu Tisch. Wie Citizen Scientists helfen, die barocke Küche zu ergründen. In *Kooperationen in Den Digitalen Geisteswissenschaften Gestalten. Herausforderungen. Erfahrungen und Perspektiven*; V&R Unipress: Göttingen, Germany, 2020; pp. 127–140.
36. Kensing, F.; Blomberg, J. Participatory design: Issues and concerns. *Comput. Support. Coop. Work (CSCW)* **1998**, *7*, 167–185. [[CrossRef](#)]
37. Benito-Santos, A.; Dorn, A.; Gómez, A.G.L.; Palfinger, T.; Sánchez, R.T.; Wandl-Vogt, E. Playing Design: A Case Study on Applying Gamification to Construct a Serious Game with Youngsters at Social Risk. *J. Comput. Cult. Herit.* **2021**, *14*, 13:1–13:19. [[CrossRef](#)]
38. Ericsson, K.A.; Simon, H.A. Verbal Reports as Data. *Psychol. Rev.* **1980**, *87*, 215–251. [[CrossRef](#)]
39. Zahidi, Z.; Lim, Y.P.; Woods, P.C. Understanding the user experience (UX) factors that influence user satisfaction in digital culture heritage online collections for non-expert users. In Proceedings of the 2014 Science and Information Conference, London, UK, 27–29 August 2014; pp. 57–63.
40. Armaselu, F.; Jones, C. Towards a Digital Hermeneutics? Interpreting the User’s Response to a Visualisation Platform for Historical Documents. In Proceedings of the DH Benelux Conference 2016, Luxembourg, 9–10 June 2016.
41. Lam, H.; Bertini, E.; Isenberg, P.; Plaisant, C.; Carpendale, S. Empirical Studies in Information Visualization: Seven Scenarios. *IEEE Trans. Vis. Comput. Graph.* **2012**, *18*, 1520–1536. [[CrossRef](#)] [[PubMed](#)]
42. Isenberg, T.; Isenberg, P.; Chen, J.; Sedlmair, M.; Möller, T. A Systematic Review on the Practice of Evaluating Visualization. *IEEE Trans. Vis. Comput. Graph.* **2013**, *19*, 2818–2827. [[CrossRef](#)]
43. Correll, M.A.; Alexander, E.C.; Gleicher, M. Quantity Estimation in Visualizations of Tagged Text. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Paris, France, 27 April–2 May 2013; Association for Computing Machinery: Paris, France, 2013; pp. 2697–2706. [[CrossRef](#)]
44. Bouckaert, R.R. Bayesian Belief Networks: From Construction to Inference. Ph.D. Thesis, Utrecht University, Utrecht, The Netherlands, 1995.